



## 30차시 다문화교육과 평가

### 1. 평가의 원리

#### 1) 평가의 기원

평가는 우리 삶의 모든 영역에서 중요한 역할을 한다. 넓은 의미에서 보면 아침에 날씨를 평가하고 무슨 옷을 입을지 결정하고, 내 몸 상태를 평가하고 아침을 무엇을 어느 정도 먹을지 평가한다. 출근시간, 직장간의 거리, 피곤함, 교통체증 등을 종합 평가해서 교통편을 선택한다. 아침부터 시작된 우리 삶의 모든 영역이 평가의 연속이다. 그러나 우리가 좁은 의미의 평가를 생각할 때는 주로 학업성취평가를 생각한다. 대한민국 국민이라면 학교에서의 평가가 내 삶에 얼마나 큰 영향을 주었는지 알 것이다. 특히 교육분야에서 평가의 역할은 일상화되어있다.

평가가 우리 삶에서 이렇게 큰 역할을 하게 된 것은 그리 오랜 역사를 가지고 있지 않다(배호순, 1991). 평가의 개념이 처음 도입된 것은 산업혁명이후 영국의 복지프로그램이었다. 산업혁명으로 도시에 몰려든 노동자들의 삶이 열악해지고, 특히 부녀자의 삶이 비참하였다. 이에 영국 정부는 처음으로 복지프로그램을 수행했는데 그 효과가 궁금해서 프로그램의 효과를 알아보려는 노력을 했다. 이것이 계기가 되어 평가에 대한 관심이 일기 시작했고, 이를 계기로 평가는 여러 영역에 사용되기 시작했는데 학교교육에서 가장 활발하게 사용되고 연구 발전되었다.

처음에는 인간의 능력을 어떻게 수량화할 수 있는지에 대한 쟁론과 이를 해결하려는 노력이 평가의 주된 관심이었다. 가장 대표적인 노력이 지능(IQ)의 측정이었다. 인간의 능력을 절대적 또는 상대적으로 수치화하는 것이 가능하게 된 것이다. 물론 초창기는 그 신뢰도가 낮았지만 그것은 연구와 실험을 통해 높일 수 있음을 보여주었다. 이후 다른 다양한 영역에서도 평가가 중요한 역할을 하기 시작하였다. 적성검사, 기질검사, 창의성검사, 도덕판단능력검사, 역량검사 등 개인평가는 물론이고, 기관 평가에 이르기까지 평가는 모든 분야에서 중요한 활동이 되었다.

## 2) 평가의 관점

평가가 우리의 삶에서 중요해지면서 평가를 어떻게 볼 것인가에 대한 관점이 다양해지기 시작했다. 평가관은 다음과 같은 순서로 변해왔다.

### 1) 측정관(measurement)

처음 평가하기 시작했을 때는 과연 인간의 능력을 수치로 측정이 가능한가라는 논쟁이 일어났다. 인간의 능력을 객관적으로 표현하지 못한다면 평가의 기능은 그 만큼 축소될 수밖에 없기 때문이다. 그러나 이것이 수량화된다면 그 활용도는 엄청나게 커진다. 그래서 초창기는 인간의 능력을 어떻게 정확하게 수량화할 것인가에 초점이 맞추어졌다. 예를 들어 “A학생은 B학생보다 머리가 좋다.”라는 표현보다 “A학생이 지능이 130이고, B학생의 지능이 120이다.”라는 표현이 훨씬 활용도가 높을 것이다. 후자의 경우는 어느 정도 차이가 나는지 정확하게 비교할 수 있을 뿐만 아니라 수많은 학생들의 지능도 비교할 수 있다. 또한 이러한 학생의 능력을 수치화해서 우수한 학생을 선발하고 배치하는 데에도 활용할 수 있다. 다만 이런 점수가 얼마나 정확한가 하는 것이 숙제로 남는다. 그러나 과학자들은 연구와 실험을 통해 신뢰도는 계속 높아왔고 이러한 노력이 평가 연구의 발전과 확산에 기여하였다. 대표적인 것이 IQ의 등장이었다. 지금도 IQ는 가장 많이 사용되는 인간 능력의 표현이다. 인간의 능력뿐만 아니라 정책 프로그램의 평가도 마찬가지이다. 정확한 수치화가 가능해야 정책효과의 크기와 비교가 가능하다.

### 2) 평가관(evaluation)

인간의 능력이나 프로그램의 효과를 어느 정도 정확하게 측정할 수 있다는 사실에 따라 이제는 그 결과를 어떻게 활용할 것인가에 대한 고민을 하게 되었다. 단순히 학생의 능력에 따른 선발 배치 또는 정책 프로그램의 효과 확인 및 개선을 넘어 좀 더 그 수치를 다양하고 유용하게 활용할 수 있을까 하는 것이다. 즉 단순한 측정의 수준을 넘어 더 발전된 활용의 방안을 연구하게 되었다. 예컨대 수업목표를 학생들이 얼마나 달성했는지 확인하고, 달성하지 못한 부분을 해결하기 위한 개선의 자료를 사용한다든가, 정책 프로그램의 어떤 부분은 효과적이었고 어떤 부분은 효과가 없었는지를 확인해서 개선의 자료로 쓸 수도 있을 것이다. 즉 대상의 측정을 넘어 그 결과를 어떻게 활용할 것인가에 대한 관점을 평가관이라고 한다.

### (3) 총평관(assessment)

평가의 격언 중에 ‘평가되지 않으면 성취하지 못한 것(not find it, not get it)’이라는 표현이 있다. 교사가 학생의 능력을 평가해내지 못하면 그 학생은 그 능력이 없는 것이다. 왜냐하면 학생은 스스로 자신이 어떤 능력을 가지고 있는지 잘 알지 못하기 때문에 그것을 꺼내어 발전시킬 기회를 상실하기 때문이다. 교사나 부모가 그 학생이나 자녀의 능력을 찾아내어 아이에게 인식시키는 것이 중요하다는 사실을 말해 준다. 그래서 어릴 때 현명한 선생님과 현명한 부모를 만나는 것이 중요하다. 총평관은 이러한 철학에 기초해 있다. 마치 의사가 환자를 진료해서 병명을 정확하게 찾아내어야 그에 따른 정확한 처방이 가능하다는 논리와 같다. 즉 정확한 진단이 중요한 것이다. 그래서 처방적 평가관이라고도 한다. 이전의 평가관이 주로 집단에 대한 관심을 가졌지만 총평관은 개인 한 사람에게 관심을 가진다. 그 개인의 능력을 정확하게 평가하는 데에 초점을 둔다. 그래서 수행평가의 개념이 등장하였다.

수행평가란 평가의 타당도와 정확도를 높이기 위해 전통적인 지필평가와 같이 간접적인 방식

이 아니라 학생의 직접 수행으로 능력을 평가하는 것이다. 좀 더 구체적으로 정의하면, 자신의 지식이나 기능을 나타낼 수 있도록 산출물을 만들거나 행동으로 보여주어서 답을 작성하는 방식이다(백순근, 2008). 최근 역량중심 교육과정의 강조되면서 역량을 평가하는 움직임이 더 강조되었다. 역량이란 할 줄 아는 능력을 말한다. 예를 들어 문제해결역량이란 문제가 주어졌을 때 그 문제를 얼마나 잘 해결하는지가 역량이 될 것이다. 평가도 문제해결과정을 관찰을 통한 기술로 평가는 것이 강조된다. 그러한 기술이 진로상담이나 입시 등에 많이 반영될 것으로 예상된다.

현실적으로는 세 가지 관점이 모두 사용되고 있다. 그러나 무게중심은 총평관으로 옮겨가고 있다. 왜냐하면 측정관이나 평가관은 학생들의 상대적 위치를 결정하는 것에 치중해 있어서 학생들의 능력을 발굴하여 개발하는 적극적 의미의 평가역할을 다 할 수 없기 때문이다. 예를 들어 다른 나라 이주민의 경우 측정관이나 평가관으로 평가할 경우 한국어가 익숙지 않아 그의 능력이 과소평가될 가능성이 많다. 그러나 총평관의 관점에서는 그가 가진 능력을 최대한 평가해내고 능력을 발휘할 수 있도록 처방이 가능할 것이다. 다문화교육에서 평가는 총평관의 입장에서 사용되어야 바람직하다.

### 3) 평가의 조건

제대로 된 평가가 되기 위해서는 두 가지 조건이 필요하다.

#### 1) 타당도

타당도란 평가하고자 하는 것을 충실하게 평가하고 있는지를 말한다. 예를 들어 수학수업을 했는데 수학능력을 국어시험으로 평가한다면 그 평가는 타당도가 없다는 뜻이다. 한국어가 서툰 외국인에게 한글로 된 수학시험을 보게 한다면 한글을 몰라서 문제를 못 풀 수도 있는 것이다. 예를 들어 호주의 경우는 운전면허실기시험에서 여러 나라의 언어로 볼 수 있도록 배려하고 있는 것도 이 때문이다.

#### 2) 신뢰도

신뢰도는 반복하여 측정하여도 항상 같은 결과가 나오는지를 말한다. 같은 평가문항인데 실시할 때마다 점수가 다르게 나온다면 그 평가문항은 신뢰할 수가 없다는 뜻이다.

타당도	비타당도 부분	오차
-----	---------	----

신뢰도
-----

[그림] 한 검사가 측정하는 전체 영역

신뢰도와 타당도의 관계를 그림으로 나타내면 한 검사가 측정하는 전체 영역에서 항상 오차는 생기기 마련이고, 이를 제거한 것이 신뢰도이다. 그리고 신뢰도 안에서도 일정부분은 비타당도 부분도 존재하며, 이를 제거하면 타당도가 된다(황정규, 2005).

#### 4) 평가자의 자세

평가의 목적은 학생의 능력을 정확하게 찾아내자는 것이므로 평가자는 다음과 같은 자세를 가져야 한다.

##### (1) 거울식 평가

우리는 매일 거울을 본다. 몸단장이 제대로 되었는지 자기 몸에 이상은 없는지 매일 자신을 되돌아보기 위함이다. 이처럼 평가도 마치 매일 거울을 보는 것처럼 자주 실시해야 한다는 뜻이다. 의사가 환자의 상태를 자주 관찰해서 차도를 살펴보듯이 교육의 효과를 자주 관찰하여야 한다.

##### (2) 확대경식 평가

평가는 구체적이고 정교해야 한다. 교육의 효과를 자세히 관찰해야 한다는 뜻이다. 학생의 미세한 변화를 놓치지 않아야 한다. 자세히 보기 위해서는 관심을 가져야 하고, 자세히 보면 평소에 보지 못한 것들을 보게 된다. 교사가 찾아내지 않으면 학생의 능력은 없는 것과 마찬가지이다. 왜냐하면 학생은 아직 어리기 때문에 스스로 자신의 능력을 찾아내기 못하는 경우도 많기 때문이다.

##### (3) 망원경식 평가

멀리서 전체적인 전망을 가지고 평가를 해야 한다. 다문화가정 자녀가 겪는 가장 큰 어려움은 언어의 문제이다. 생활언어와 학습언어는 다르다. 생활언어로서 한국어를 익히기는 쉽지만 학습언어로서 한국어를 익히는 데에는 시간이 걸린다. 예를 들어 수학문제를 풀 수 있는데, 한국어가 익숙지 않아서 문제의 뜻을 이해하지 못해 풀지 못하는 경우도 있다. 이는 그 학생에게 한국어 문제이지 수학의 문제가 아니다. 이러한 점을 고려하지 않고 평가해버리면 학습자는 자신감을 잃어버려 자신은 무능한 사람이라고 스스로 낙인찍고 노력을 포기할 수도 있다. 그러므로 좀 더 길게 보고 기다려가며 평가하는 태도를 가질 필요가 없다.

##### (4) 광학렌즈식 평가

다양한 관점에서 평가할 필요가 있다. 다중지능이론에서 알 수 있듯이 학습자마다 다양한 능력이 있고, 다양한 적성을 가지고 있다. 이처럼 다양한 능력과 적성을 찾아내어주는 교사가 필요하다. 그래서 만남은 교육에 앞선다는 말이 있다. 즉 스승이 교육에 앞선다는 말이다. 베드로가 예수를 만나지 않았다면 초대 교황이 되지 못했을 것이다. 히딩크가 찾아내지 않았다면 박지성은 평범한 축구선수로 남았을지 모른다. 훌륭한 교사는 학생의 능력을 찾아낼 줄 알아야 한다. 그러기 위해서는 다양한 렌즈로 학습자를 평가할 수 있어야 한다.

## 2. 평가의 절차

### 1) 수업목표와 평가목표의 확인

평가를 하기 위해 제일 먼저 해야 할 일은 수업목표의 확인이다. 그래야 어떤 내용을 어떤 방법으로 평가할 것인지 방향과 범위를 잡을 수 있다. 타당도 높은 평가를 하기 위함이다. 이를 위해 교육과정의 성취기준을 먼저 확인한다. 다음 예시의 경우라면 학생들은 세계 인권문제의 양상

을 조사하고, 이에 대한 해결방안을 제시할 수 있는지를 평가해야 한다. 교과서도 성취기준을 중심으로 만들어지기 때문에 성취기준을 수업목표의 기준으로 삼는 것이 좋다.

<예시: 2015 통합사회 성취기준과 해설>

성취기준	10통사04-03 사회적 소수자 차별, 청소년의 노동권 등 국내 인권문제와 인권지수를 통해 확인할 수 있는 세계 인권문제의 양상을 조사하고 이에 대한 해결방안을 제시한다.
성취기준 해설	[10통사04-03]에서 사회적 소수자의 사례로는 장애인, 이주 외국인 등을 다룰 수 있으며, 청소년 노동권의 경우 청소년들이 일을 하면서 보장받아야 할 권리 및 관련 법규를 청소년 아르바이트와 같은 사례에 적용하여 다룰 수 있다. 세계 인권 문제는 국제기구나 비정부 기구 등에서 발표하는 인권지수를 활용하여 세계 각 지역에서 나타나는 인권 문제의 양상과 해결 방안을 다루도록 한다.

다음으로는 평가목표의 설정이 필요하다. 수업목표가 달성되었는지 모두 다 평가하기에는 한계가 있다. 그러므로 교사의 전문적인 선택이 필요한데, 어떤 내용을 어떤 방법으로 무엇을 평가할지 평가 목표를 세울 필요가 있다. 학생의 성취여부를 확인할 대표적 평가문항을 만들어야 한다. 그리고 어느 정도의 성취를 보여주면 만족할 수 있을지도 생각해 두어야 한다. 또는 더 나아가 수업을 제대로 했는지를 알아보기 위한 평가도 고려할 필요가 있다.

## 2) 평가영역과 내용의 선정

평가영역이란 소위 ‘평가 범위’를 말한다. 학생들은 시험범위라고도 한다. 평가 영역을 정한 다음에 평가에 사용할 내용을 정해야 한다. 평가내용은 교육과정의 성취기준에 근거하면 된다. 다만 내용과 관련하여 구체적으로 학생의 어떤 행동을 평가할 것이냐를 결정할 때는 행동과 내용의 이원분류표 작성하는 것이 도움이 된다. 이원분류표란 주로 내용 영역과 행동 영역을 구분하여 표로 만든 것이다. 즉 학생에게서 나타나야 할 행동, 그 행동과 관련된 내용의 두 가지 요소를 표현한다. 이렇게 하는 이유는 대부분의 평가가 단순한 지식 암기만을 확인하는 경향으로 흐르기 때문에 이를 방지하기 위함이다. 광학렌즈식 평가를 강조한 것처럼 학생들의 다양한 사고력을 평가하기 위해 평가문항이 이원분류표에 골고루 분포되도록 하기 위함이다. <표 13-1>의 예처럼 ‘사회적 소수자의 개념’에서는 이해를 평가하고, ‘외국인 근로자’에서는 적용을, ‘장애인’에서는 종합의 행동 능력을 평가한다는 뜻이다.

<표> 이원 분류표

내용 \ 행동	지식	이해	적용	분석	종합	평가
1. 사회적 소수자의 개념		○				
2. 외국인 근로자			○			
3. 장애인					○	

## 3) 평가방법의 결정

평가방법은 기준에 따라 다양하게 분류된다.

첫째, 지필평가와 구술평가이다.

대부분의 평가는 지필평가로 이루어지지만 최근 구술평가의 중요성이 많이 강조되고 있고, 실제 입시에서도 면접의 중요성이 점점 커지고 있다. 우리는 실제 생활에서 글보다는 말을 더 많이 한다. 말로 자신의 의견을 표현하는 것이 매우 중요하다는 뜻이다.

둘째, 전통평가와 수행평가이다.

전통적인 지필평가에 대한 반성으로 수행평가가 많이 강조되고 있다. 왜냐하면 학생의 능력을 정확히 평가하기 위해서는 학생이 직접 수행하는 것을 평가하는 것이 타당하기 때문이다. 두 평가의 차이점은 다음과 같다.

<표> 전통 평가와 수행 평가의 비교

구분	전통 평가	수행 평가
학습관	추상적, 객관적 상황	구체적, 주관적 상황
학습관	기억과 재생산	이해와 성장
평가 체제	상대평가와 양적 평가	절대평가와 질적 평가
평가 내용	내용 지식(결과)	실천 지식(과정과 결과)
평가 방법	선택형 표준검사	다양한 교사 평가
평가 시기	학습종료 후 평가	모든 학습과정
학생의 역할	수동적, 지식의 재생산	능동적, 지식의 창조자
수업 활동	교사 중심	학생 중심

자료: 백순근(2000). 수행평가의 원리. 서울: 교육과학사, pp. 53-54.에서 수정

셋째, 개별평가와 모둠평가이다.

우리의 삶은 대부분 여러 사람이 더불어 살아가는 것이고, 지식이 전문화되고 세분화되어 서로 협력없이 문제를 해결하기 힘든 사회가 되었다. 따라서 다른 사람들과 얼마나 잘 협력을 하느냐가 중요한 능력으로 인정되고 있다. 그러므로 개별평가 외에 모둠평가가 강조되고 있다. 모둠을 평가하는 방식은 다음과 같이 다양하며 각각 장단점이 있다.

<표> 모둠 평가의 방법과 특징

방법	특징
모둠원 점수를 단순 합산한 평균점수	간단한 방법이지만 모둠원 개인의 부담이 큼
모둠원이 함께 만든 모둠작품 점수	모둠원 개인이 어떤 기여를 했는지 알기가 힘들
모둠원이 모두 통과되어야 모둠점수 부여	모둠원 개인의 부담감은 있지만 다른 모둠원이 도와주도록 하는 효과가 있음
임의의 모둠원 한 명을 평가하여 모둠점수 부여	누가 선정될지 모르기 때문에 모둠원이 열심히 참여하고 서로 도와주지만 개인의 부담감이 큼
개인의 과거 점수에서 얼마나 향상되었는지를 계산해서 향상점수를 단순 합산한 평균 점수	각자 개인의 과거와의 경쟁이므로 모둠원들이 협력하고 열심히 노력함

#### 4) 평가 시점

평가 시점에 따라 세 가지 평가방법이 있다.

첫째, 진단평가는 출발점 평가라고 하는데, 수업을 하기 전에 학생들의 수준을 진단하기 위해 평가를 하고, 이 결과를 보고 수업의 수준과 속도 등을 결정하는 데에 활용한다.

둘째, 형성평가는 한 단위 또는 한 차시 수업도중 현재 수업목표를 잘 달성하고 있는지 확인하기 위해 현재까지 배운 내용을 평가하는 것이다. 그 결과를 보고 그 동안의 수업에 대해 반성하고, 남은 수업시간동안 학습목표 달성 전략을 수립한다.

셋째, 총괄평가는 한 단위 또는 한 차시 수업이 끝난 후 수업목표 달성여부를 총괄적으로 평가한다.

#### 5) 평가결과의 해석

평가결과 해석과 관련하여 두 가지 방식이 있다. 첫째는 학생의 성취를 평가하는 것이다. 학생을 평가할 때는 두 가지 방식이 있다. 하나는 절대평가로 특정 준거(criterion)를 기준으로 하여 학생의 수준을 평가하는 방식이고, 다른 하나는 상대평가로 규준(norm)의 분포에서 학생이 상대적으로 어떤 위치를 차지하는지를 기준으로 평가하는 방식이다. 여기서 규준이란 좌우대칭의 정상분포 곡선을 말하는데, 한 학생의 상대적 위치를 알 수 있다.

둘째는 평가결과를 수업의 개선에 활용하는 것이다. 즉 평가의 목적이 학생이 대상이 아니라 수업이 대상이다. 만약 수업목표가 달성되지 않았다면 어떤 부분에서 문제가 있었는지 평가를 통해 알 수 있다. 학생의 학업성취결과에서 수업의 문제점을 찾아낼 수도 있고, 별도로 수업에 대한 평가를 활용할 수도 있다.

### 3. 평가문항 개발의 원리

#### 1) 평가 문항의 종류

평가를 하기 위해서는 평가문항을 만들어야 한다. 평가문항을 어떻게 만드느냐에 따라 평가결과의 유용성이 결정된다. 보통 평가는 일련의 평가문항의 세트로 이루어진다. 평가문항의 종류는 여러 가지 기준에 따라 다양하게 분류할 수 있다.

첫째, 평가자의 기준으로 나눌 수 있다.

평가자 즉 채점자의 판단에 따라 점수가 달라진다면 이는 채점자의 주관에 개입된 경우이다. 이런 경우 주관식 평가라고 한다. 반면에 채점자가 누구이든 아무리 많은 평가 결과가 똑 같이 나온다면 이는 객관식 평가라고 한다(황정규, 2005).

둘째, 수험자의 기준으로 나눌 수 있다.

<표> 평가문항의 종류(수험자의 기준)

종류	유형		
선택형(객관식)	진위형	선다형	배합형
서답형(주관식)	단답형	서술형	논술형

### (1) 선택형

선택형은 수험자가 주어진 선택지에서 정답을 고르는 형식인데, 세 가지 유형이 있다.

첫째, 진위형은 흔히 OX문항이라 불리는 것으로 주어진 지문에 대해 '진실'과 '거짓'으로 표현하게 한다. 일견 출제나 응시나 쉬운 것 같으나 명백한 진위가 있는 경우가 없으므로 문제내기도 어렵고 맞히기도 어렵다.

둘째, 선다형은 주어진 답지가 몇 개인지에 따라 4지선다, 5지선다 등으로 부른다. 주어진 답지 중에 정답을 고르는 방식이다. 때로는 복수로 고르게 하는 문항도 사용한다.

셋째, 배합형은 A군의 답지와 B군의 답지를 연결하게 하는 방식이다. 직접 연결하게도 하고, 보기를 2개 이상 주고 각각 선택하게도 한다.

### (2) 서답형

서답형은 수험자가 자신의 대답을 글로 쓰는 형식이다. 단답형과 서술형은 주로 외워서 쓰는 경우가 많고, 논술형은 비판적 사고를 보이면서 자신의 주장을 논증과 실증으로 정당화하는 방식으로 쓴다.

첫째, 단답형은 정답이 어휘의 형식으로 되어서 직접 묻는 방식과 문장 중에 빈칸을 주고 적게 하는 방식이 있다.

둘째, 서술형은 정답을 문장으로 표현하여 답하게 하는 방식이다.

셋째, 논술형은 주어진 주제에 대해 논증이나 실증으로 자신의 주장을 쓰게 하는 방식이다.

### (3) 질문 형식의 기준

평가문항의 문두에서 어떤 지시를 하느냐에 따라 여러 가지 유형이 있다.

#### ①최선답형(best-answer variety)

최선답형은 여러 답지 중에 가장 맞는 답, 정답의 정도가 가장 큰 것을 선택하게 하는 방법인데 주로 '가장'이라는 표현을 많이 사용한다. 복수 정답시비를 없앨 수 있는 좋은 방법이다.

#### ②정답형(correct-answer variety)

한 개만 정답이고 다른 것은 오답으로 답지를 구성하는데, 오답을 만드는 것이 힘들고, 오답시비가 나타날 가능성이 있다. 명백히 맞고 틀려야 하기 때문에 사실 확인 문항이 많고, 사고력을 요하는 문항을 만들기가 어렵다.

#### ③합답형(combined-response variety)

합답형은 여러 개의 답 중에서 둘 이상 합해서 정답이 되는 형식이다. 합해진 답지 세트를 고르게 하는데 생각을 많이 하게 하므로 문항의 난이도를 높일 때 많이 사용한다.

#### ④부정형과 긍정형(negative variety)

문두가 '아닌 것은?', '거리가 먼 것은?' 등 부정형으로 물을 수도 있고, 긍정형으로 물을 수도 있다. 부정형으로 할 때는 반드시 밑줄을 그어야 하는데 이는 수험자가 실수하지 않도록 하기 위함이다. 가능하면 긍정형으로 문항을 사용하라는 요구가 있는 것은 이 때문이다. 그러나 수험생이 오해하거나 실수할 염려가 없다면 부정형이든 긍정형이든 크게 문제되지 않는다.



## 2) 평가 문항의 제작원리

### (1) 평가 문항의 개념과 준비

문항이란 채점단위(scoring unit)를 말하며 이의 조합이 검사 또는 시험문제이다. 문항의 제작은 오답시비가 없고, 배운 내용을 담아서 만들어야 하는 고도의 창의적 활동이다. 문항제작자는 다음과 같은 준비를 해야 한다.

첫째, 평가내용에 대한 충분한 지식이 있어야 원하는 대로 문항을 만들 수 있다. 왜냐하면 한 문항에는 표현된 것 외에 보이지 않는 많은 의미들을 담아야 하기 때문이다.

둘째, 수업목표를 확인하고, 다음으로 평가목표를 확인하는데 이것이 주요한 것은 타당한 평가문항을 만들기 위함이다. 수업목표와 평가목표가 반드시 일치하는 것은 아니다. 수업목표를 모두 평가하는 것은 불가능하다. 평가 시간을 확보하기 힘들고 평가할 수 있는 영역도 제한되어 있기 때문이다. 그러므로 학생의 능력을 효율적으로 평가할 수 있는 최소한의 평가목표를 별도로 세우는 것이 현명하다. 평가문항을 만드는 과정에서 계속 수업목표와 평가목표를 의식하고 확인하면서 진행해야 한다.

셋째, 학생의 심리발달을 이해해야 한다. 이는 난이도와 관계가 깊다. 학생의 수준을 잘 알고 있어야 적절한 수준의 문항을 만들 수 있다.

넷째, 지필검사인 경우 문항제작에 필요한 언어적 능력이 있어야 한다. 즉 학생들과 의사소통이 되는 문항을 만들어야 한다는 뜻이다. 묻고자 하는 것이 무엇인지 학생들이 쉽게 이해되는 표현을 사용해야 한다. 문제가 어려운 것이 아니라 문두, 자료, 답지의 표현 자체가 어렵다면 이는 좋은 문항이 아니다. 왜냐하면 문항을 잘못 해석하거나 문항해석에 너무 많은 시간을 사용해서 정작 평가하고자 했던 학생의 능력을 평가해내지 못할 수 있기 때문이다.

다섯째, 문항의 종류와 제작 방법을 잘 알아야 한다. 문항의 형식적 종류는 무척 많고, 문항 제작은 창의적인 활동이다. 그러므로 사전에 문항에는 어떤 종류가 있고, 각 문항의 장단점을 잘 알고 있어야 한다. 평가문항은 '눈금자(scale)'와 같다. 눈금자가 정확해야 길이를 정확하게 재는 것처럼 평가문항은 학생의 능력을 확인하는 것이므로 오류가 없어야 한다.

### (2) 문항제작의 점검과 조건

문항을 제작할 때는 다음을 점검해야 한다.

첫째, 평가시기를 결정한다. 전술한 바와 같이 진단검사, 형성검사, 총괄검사에 따라 시기가 다르다.

둘째, 평가 문항의 유형을 결정한다. 평가 목표에 맞게 유형들을 결정한다.

셋째, 수업 목표와 부합하는지 확인해야 한다. 그래야 타당도가 높아진다.

넷째, 문항수를 결정한다. 문항수가 지나치게 많거나 적어도 문제가 생긴다. 주어진 시험시간을 고려해서 적절한 문항수를 만든다.

문항은 대부분 문두-자료-답지의 형식을 가지고 있다. 형식적으로 좋은 문항이 되기 위해서는 다음과 같은 조건이 필요하다.

- 학생이 배운 내용이 적절히 반영된 장면이어야 함
- 문항장면이 적절히 복합적이어서 일상생활과 연결되어야 함
- 지나친 추상적 표현은 지양해야 함
- 문법적으로 정확해야 함
- 학생이 실수로 답하지 않도록 구조화 함
- 내용과 관계없이 형식으로 정답을 유추할 수 있으면 안 됨

- 지시문이 분명해서 애매하지 않아야 함
- 지시문이 길거나 복잡적이어서 학생들이 잘못 이해하지 않아야 함
- 지시문에서 강조할 부분, 조심해야 할 부분을 밑줄 등으로 표시함
- 문항유형의 특징과 평가목표가 적합해야 함